Governing artificial intelligence in the public interest

Mariana Mazzucato
Professor in the Economics of Innovation and Public Value
Founding Director, UCL Institute for Innovation and Public Purpose

Marietje Schaake
International Policy Director
Cyber Policy Center, Stanford University

Seb Krier
Senior Technology Policy Researcher
Cyber Policy Center, Stanford University

Josh Entsminger
PhD student in innovation and public policy
UCL Institute for Innovation and Public Purpose

WORKING PAPER
WP 2022/12

UCL Institute for Innovation and Public Purpose

Stanford | Cyber Policy Center
Freeman Spogli Institute and Stanford Law School

VISIT…

LANZAROTE Caliente.com

About the Stanford Cyber Policy Center

The Stanford Cyber Policy Center, a joint initiative of the Freeman Spogli Institute for International Studies and Stanford Law School, is Stanford University's research center for the interdisciplinary study of issues at the nexus of technology, governance and public policy focused on digital technologies impacting democracy, security, and geopolitics globally. Led by Nathaniel Persily, the James B. McClatchy Professor of Law at Stanford Law School and Dan Boneh, the Rajeev Motwani Professor in the Stanford School of Engineering, the center connects academia, the legal and tech industry and civil society with policymakers around the country to address the most pressing cyber policy concerns.

Through research, policy engagement, and teaching, the Cyber Policy Center aims to bring cutting-edge solutions to national governments, international institutions, and industry. Working across disciplines including law, the center's six programs are staffed with world-renowned researchers and academics who aim to understand how digital technologies are impacting democracy, security, and geopolitics globally.

About the Institute for Innovation and Public Purpose

The UCL Institute for Innovation and Public Purpose (IIPP) aims to develop a new framework for creating, nurturing and evaluating public value in order to achieve economic growth that is more innovation-led, inclusive and sustainable.

We intend this framework to inform the debate about the direction of economic growth and the use of mission-oriented policies to confront social and technological problems. Our work will feed into innovation and industrial policy, financial reform, institutional change, and sustainable development.

A key pillar of IIPP's research is its understanding of markets as outcomes of the interactions between different actors. In this context, public policy should not be seen as simply fixing market failures but also as actively shaping and co-creating markets. Re-focusing and designing public organisations around mission-led, public purpose aims will help tackle the grand challenges facing the 21st century.

IIPP is housed in The Bartlett, a leading global Faculty of the Built Environment at University College London (UCL), with its radical thinking about space, design and sustainability.

This IIPP Working Paper Series is published electronically at:

https://www.ucl.ac.uk/bartlett/public-purpose/publications/working-papers

ISSN 2635-0122

This report can be referenced as follows:

Mazzucato, M., Schaake, M., Krier, S. and Entsminger, J. (2022). *Governing artificial intelligence in the public interest*. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2022-12). Accessed: https://www.ucl.ac.uk/bartlett/public-purpose/wp2022-12

Governing artificial intelligence in the public interest

Mariana Mazzucato*, Marietje Schaake**, Seb Krier** and Josh Entsminger*

Abstract

Artificial intelligence (AI) is being positioned as the defining general purpose technology of the early 21st century. Nations are racing to secure leadership in AI innovation capabilities, focusing on horizontal innovation policy to drive domestic technological competitiveness. However, we argue global and national administrations need to be more focused on the direction of AI innovation. We propose that a market-shaping approach can help align public and private interactions to drive AI towards advancing public interests. We further propose that this requires a bolder global technology policy agenda to align and shape State and non-State actors' relationships within AI development and diffusion towards public interest activities. We conclude with a series of limited recommendations to help policymakers frame a market-shaping approach to AI governance.

Keywords: public interest, public purpose, artificial intelligence, United States, market shaping, regulation, technology policy

JEL codes: A13, F50, H10, H23, H40, H83, H87, I30, I38, L10

Acknowledgements:

This research was built on work funded by the Omidyar Foundation and the Rockefeller Foundation.

* UCL Institute for Innovation and Public Purpose
** Stanford University, Cyber Policy Center

1. Introduction

Artificial intelligence (AI) has emerged as a defining technology of the early 21st century (Hogarth 2019). Historically, the United States government played a leading role in funding research into novel AI solutions and investing in the infrastructure to bring such solutions to the market (OTA 1999). Major technology corporations continue to reap the benefits of public investment that started in the 1950s and continues until today. Early investment into computing capacity, for example, has laid the groundwork for a new slate of powerful AI tools (Fuchs 2010). Given the growing uptake of these AI technologies, we must now focus on assessing the potential systemic harms from their under-governed application. Some problems have already surfaced and more must be prevented. Specifically, the unregulated expansion of AI may exacerbate economic inequality and violate human rights.

In this paper, we argue that the US should adopt a new approach for developing and shaping AI capabilities by integrating the technologies in policies specifically geared towards a sustainable, equitable and green digital economy. A singular focus on competition and innovation is myopic. Policymakers must consider the desired outcomes for society writ large to both avoid long-run failures in the development of AI capabilities and adequately address the grand challenges of our time. More directly, a focus on AI competitiveness and US industry leadership must address vulnerabilities in the existing AI innovation system. Competition-focused strategies could tend towards value-extractive economic futures that will lead to important socio-economic harms, such as the concentration of market power and negative effects on labour markets.

We propose policies that shape US markets while aligning innovation and competition policies towards creating a domestic AI innovation system that better serves the public interest. Such policies should aim to better align domestic investment and AI capability development with economic, societal and national security objectives. Without such early tuning, AI investments can lock in sub-standard implementations and socio-economic pathways, creating long-run issues which the state will have to address sooner or later.

We propose a market-shaping perspective as a novel, alternative analytic framework for repositioning the relationship between the public and private sector in the future of AI markets for the US. Existing market failure-based approaches to governing the innovation process, while crucial, may insufficiently address the intersection between innovation and competition policy. Furthermore, the scope of potential advantages and disadvantages from AI benefits from a pre-distributive focus concentrating on prevention instead of later reactive cures. The risk of path dependence needs to be thought about early, to avoid fostering a suboptimal technological ecosystem. To ensure that AI serves the long-term common interest in addition to short-term corporate interests, the public and private sectors must rethink how they relate to shape the next generation of the US digital economic transition.

2. Public interest

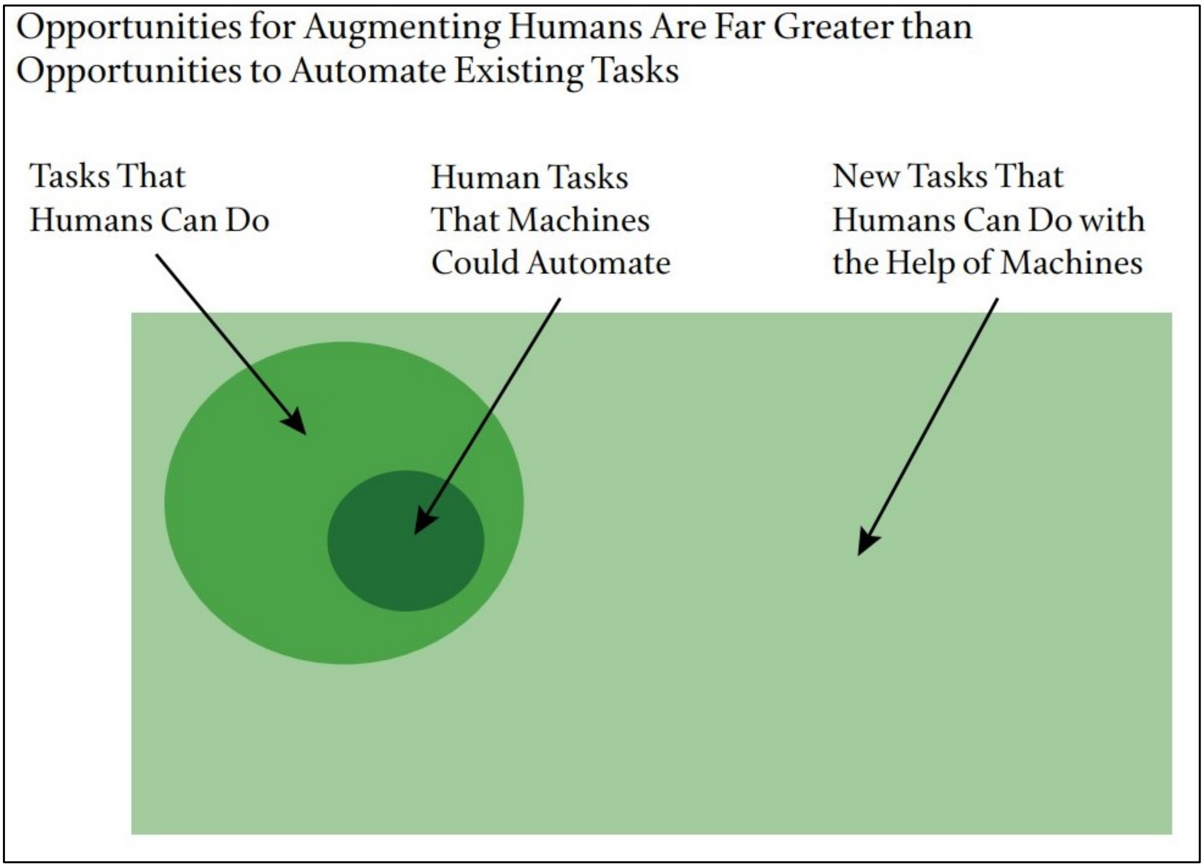
Academic researchers and corporate actors have endowed AI with a bold mission: to replicate, automate and exceed human level intelligence and capabilities. Throughout history, AI has encompassed a set of technologies with approaches and applications in a variety of sub-areas, including human intelligence, computer vision, machine translation, pattern recognition and natural language processing. This diversity poses a definition challenge, which has implications for both regulation and research. Legal and political institutions need a stable definition of AI to effectively govern the technology. Similarly, communities seeking to organise AI as a coherent research field must first reach a consensus of terms (Bryson 2022) defines intelligence within AI as 'computing action from context' to provide a more holistic and horizontal definition of intelligence for legal and regulatory frameworks). As such, it is difficult to precisely define the many components that make up the AI R&D ecosystem, but there are, at the very least, combinations of technical capabilities, hardware improvements, data creation, data management, curated datasets, machine resource management, machine learning operations management, institutional support and deployed models. Beyond definitions, the constituent parts of AI systems play a major role in determining systems. In fact, the history of AI innovation illustrates the importance of an expanding access to available and cheap compute power as essential for enabling the democratisation of AI innovation. However, as the means of accessing compute power are increasingly organised through online platform actors along with knowledge resources (such as Kaggle and Github), the conditions of AI's democratisation are increasingly being privatised.

Ultimately, AI differs from most innovations, because of its general-purpose nature. Applications range from using satellite imagery to estimating population levels to translating natural language to code. Machine learning (a subfield of AI) can accelerate the achievement of the UN's Sustainable Development Goals through smart electric grids and AI-powered healthcare. Automation has been particularly helpful in outsourcing dull or dangerous tasks; computers can independently complete certain compliance tasks in banking regulation and robots can help navigate pipeline inspections. But while markets tend to do a good job at unearthing and scaling commercial applications, some public value remains untapped, and the costs of risk and harm are offloaded to society.

2.1 Direction of automation

Erik Brynjolfsson argues that when AI *augments* human capabilities, enabling people to do things they could not do before, then humans and machines are complements (2022). Whereas when full automation leads to human labour substitution, workers lose economic and political bargaining power (Acemoglu 2020). He also explains that the scope of potential innovations is arguably higher when both humans and robots work together, since *automation* can only automate what humans can already do, whereas *augmentation* offers a much wider range of possibilities.

Figure 1: The Turing trap: the promise and peril of human-like artificial intelligence



Source: Erik Brynjolfsson (2022): The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence, p.279.

How automation is deployed in a society depends on policies such as tax incentives, liability and responsibility rulings, and corporate governance frameworks enacted by the state, rather than simply the technical capacity of the system to outperform a human at a given task (Acemoglu 2020). Daron Acemoglu also suggests that existing fiscal policies skew incentives towards excessive automation, while a more proactive approach would allow governments to optimally tilt the nature of innovation towards more human-friendly technologies (Acemoglu 2021). Acemoglu offers the example of renewable tech: government subsidies, advance purchases and changing consumer norms have allowed this technological future to flourish. Inaction, on the other hand, would have likely cemented our dependency on fossil fuels. The same can be achieved with emerging technologies such as AI: governments should in principle be able to redirect efforts and support in a socially desirable direction (Mazzucato et al 2020).

2.2 Malicious and harmful uses of AI

Without clear public policies to steer AI and other new technologies, technical and corporate standards end up setting *de facto* norms with great societal consequences. While some new technologies may be harmless, others may cause severe damage. One example is emotion recognition. Some applications can be completely harmless, such as evaluating a classroom's engagement level. but others, including use by law enforcement to read individuals, are

problematic both in terms of accuracy and overstated capabilities. Computer vision may be used to find missing children and detect illegal tree logging, but it can also be used thoughtlessly — and even maliciously — by law enforcement and non-state actors (see Wakefield 2021). Over time, the flawed application of these technologies can cause potentially undesirable disruptions of democratic norms due process and the presumption of innocence.

While the technological frontier for AI capabilities is advancing, the institutional and organisational models for public oversight — evaluating developments and enforcing the opening or closing of different pathways — of AI usage are lagging behind. One reason is the asymmetry in access to information between companies and public institutions. The proprietary nature of most AI applications means the public lacks insight as well as the ability to design proper oversight. Advancing technical capabilities without matching adjustments to governance, institutional and organisational models is leading to failure in effectively evaluating the risks of AI and managing its opportunities.

How a particular AI technology is used can vary greatly depending on context. Legal protections are always context-specific, meaning the same technology can have widely varying impacts on people. The creation of a desirable context deserves more consideration from the government: should AI's *use* be banned or is it possible to devise regulatory frameworks, policies and safeguards to promote one context, and not the other? Moreover, can governments encourage certain innovation pathways to discourage unnecessarily broad surveillance, and encourage more prosocial uses of the technology? We contend that this is both possible and desirable. Government can play a transformative role in shaping both how AI is used and strengthening the institutions by which different uses can be identified, monitored, and shut down at scale.

2.3 Directionality and societal objectives

Strategies that prioritise innovation and investment have traditionally focused on improving the *rate* of technological development, but the *direction* of that development has received less attention. While improvements to the rate of AI innovation have promise for expanding horizontal AI capabilities, the direction of innovation highlights how some technological trajectories come to dominate. Policymakers can track these trajectories and leverage them to organise social and democratic agency in a way that promotes desirable directions of innovation and deters undesirable methods (Mazzucato 2013; Stirling 2008). Having a point on the horizon matters: the pursuit of a low-carbon transition, for example, allows for the comparative assessment of the maturity, risks and advantages among different green energy mixes (Stirling 2015). Even though various option outputs are comparable in wattage, the goal of low carbon forces officials to carefully analyse their externalities. For instance, they must consider the end-of-life decommissioning costs for wind turbines versus nuclear reactors.

Facial recognition has become a common point in public discussions about the kinds of AI technologies that should be allowed in each society. The question for facial recognition is not simply whether it works in low- and high-risk environments, but is more directional: should it be endorsed as a desired part of the social technology stack? The broad banning of remote biometric solutions (such as facial recognition) across multiple cities testifies to a renewed work of public

and civic actors in exercising directional agency. We contend that this function can be performed both more widely and more precisely across the wide range of AI solutions available today, and in the near future.

So far, however, many technological solutions to public health threats, individualised medicine, inclusive education and climate change, among other global challenges and bold approaches, remain incomplete. Investments into AI have been framed with the promise of making horizontal improvements to how the public and private sectors can address these and other societal challenges. The hope is that AI can be positioned an omni-technology, can through careful moderation and use, may provide new tools and insights to either frame, address, or otherwise reduce grand challenges. For instance, AI promises to improve the speed, scale, accuracy, automatability and accessibility of automated translations of government services – creating improved cross-langague and cross-sensory requirement inclusion, if done appropriately. Additionally, a whole host of niche applications and areas are emerging, such as corrosion detection on boats and snow packing simulation, but the latter areas do not automatically scale to viable business models. In principle, small government agencies and other underfunded public sector organisations would be customers in these examples, but the conditions necessary to deploy these valuable technologies are lacking.

Making the comparative calculation, AI investment will naturally tend towards capital-rich environments rather than flowing towards addressing unmet social needs. Developers of AI may also face asymmetric access to datasets, as well as pressure to use available datasets, even if these may be low quality and deeply biased. The federal government can play an active role in mitigating these barriers and turning the tide of investment, but it faces a series of challenges that limits its efficacy. Simply put, the US lacks a proactive vision for AI and a robust set of policies on AI for the public good. The next section identifies the challenges that the government must overcome, along with the new paradigms that it should test.

3. Challenges

The NSCAI report on AI highlights a series of fundamental challenges to securing US AI leadership and supremacy — significant among these being the securing of domestic semiconductor manufacturing, domestic AI talent creation, diversification of AI breakthrough research investments and the creation of a technology competitiveness council to advance government leadership. While these recommendations make sense, we also want to draw attention to a set of additional challenges that slow down proactive policymaking.

Beyond challenges around talent, infrastructure and diffusion, there are also challenges to how progress is evaluated in the AI industry. There is a lack of consensus on what counts as an effective improvement to the direction of AI innovation at the most fundamental technical levels. Governments should be better equipped to understand the capabilities and impacts of deployed systems and ongoing innovations. Legislation recognises the existence of cases where commercial secrets can give way to the advancement of other values; it is unclear whether these

are too narrow or whether the status quo is optimal. Indeed, there may well be cases where democratic accountability might warrant closer scrutiny of algorithms by public authorities.

In addition, the field remains dominated by large and well-funded players, often commercial ones. Competition for talent is cut-throat and public interest organisations offer salaries that cannot compete with those of corporations. The resources necessary to innovate with AI are virtually inaccessible to anyone outside the world's biggest corporations and a handful of wealthy, elite universities. Moreover, the large datasets required to train algorithms are, for the most part, controlled by either industry or government. Academic researchers struggle to gain access to both, resulting in a loss of in-depth research, accountability and academic-led innovation. To address this, initiatives like the National AI Research Resource Task Force (funded by the Defense Authorization Act for fiscal year 2021) seek to bridge the gap and lower barriers to entry by subsidising cloud and compute access for researchers. Such initiatives are a welcome first step, but one that will prove insufficient if not complemented with a more holistic rethinking of the government's role.

The current generation of AI strategies in the US is fixated on industry leadership and competitiveness. Without more profoundly considering the public interest, systemic issues for future digital policymaking could become entrenched. Here, we review some of the challenges governments can address.

3.1 Critical resource access

AI applications require an existing problem to target, organisational culture and processes to be adapted, and access to rich and accurate datasets. Given the wide availability of data stemming from social networks and the internet more widely, it is arguably easy to develop an AI application trained on such data. However, more diverse, impactful or transformative applications for the benefit of the wider public will require datasets that are either non-existent, disparate or inaccessible. Some datasets are also incomplete. For example, density maps showing the geographical distribution of images in the ImageNet, COCO and OpenImages datasets are not representative of the world's population. Incomplete or biased datasets inevitably lead to poorer performance (DeVries et al. 2019).

The public sector can play a critical role in improving the quality and accessibility of datasets. A key example is the Human Genome Project, a 13-year-long publicly funded project initiated in 1990 with the objective of determining the DNA sequence of humans. The project incentivised the practice of wider data-sharing, and facilitated scientific research and innovation significantly (Dattani and Bechhofer 2021). Another important example is the Protein Data Bank, which provides access to 3D structure data for large biological molecules (proteins, DNA and RNA). The dataset came into existence following a grassroots campaign in the 1970s, and it perfectly illustrates the type of datasets and infrastructure public institutions can help collate to accelerate the speed of science and innovation. The Human Genome Project and the Protein Data Bank do not have to be isolated examples; the government can — and should — play a proactive role in supporting big data initiatives. However, in order to do this effectively, significant investment will

be needed to attract adequate talent and design nimbler institutions. While there are already some positive examples, they remain ad hoc rather than systematic.

Governments can indeed play a crucial coordinating role in facilitating access to publicly funded data sets without necessarily choosing winners or being overly prescriptive about a particular angle of research and development. Taiwan, for example, has taken a winner-takes-all approach by publishing all non-privacy-related data immediately after collection (Wiblin and Harris 2022). In addition to blanket publication, governments can set incentives, collect, centralise and share data sets, and reward initiatives that are necessary to realise ambitious but complex policy goals. In the UK, Transport for London's decision to share open data has enabled new products and services, while improving transparency and facilitating the shift of consumer usage from private vehicles to cycling and public transport (Deloitte 2017). Recently, Our World in Data initiated a campaign asking the International Energy Agency (IEA) to make its data available to the public. Despite being an institution that is largely publicly funded, the IEA locks the majority of its datasets behind paywalls. Following the campaign, the IEA has proposed making all its data and analysis freely available, an idea its members will vote on in the coming months (Quantum Commodity Intelligence 2022). Similarly valuable datasets exist in countless other institutions and carefully sharing them could play a major part in unlocking impactful innovations.

As the US expands the number of critical data sets that are publicly available, their conditions of use and access will become increasingly important (Bates 2014). Public sector aggregation of critical data sets may benefit from ensuring that private and civil sector access to the data, while needed for innovation, does not power or create further inequalities within the innovation system (Bates 2014; Collington 2022; Mazzucato 2018). The US needs to create a programme across various government agencies for evaluating the desirability of open data programmes and identifying the conditions for private sector access to public sector data (Collington 2022). Cases such as the development of weather-risk markets based on publicly collected meteorological data and Uber's use of public sector transportation data are illustrative.

The regulation and availability of compute power access has likewise become of critical concern. Proposals for a national research cloud provide a powerful argument on the national availability of critical resources. Computing cost has become a powerful divider between small and large-scale actors in research capabilities (Hogarth 2019, 2020; Hogarth and Beniach 2021), and computing cost access has become a key element of AI research funding requests.

Of course, opening greater access to data sets, compute and other advanced technologies can lead to problems. There are serious concerns around the long-term issue of open access, making it easier for malevolent actors to train potentially malicious systems. This could create a more open ecosystem for harmful actors to emerge. More tightly assessed and regulated conditions of access may need consideration.

3.2 The emergence of harmful and malicious uses of AI

AI tools deployed in the laboratory push the frontiers of science and save lives, but certain AI tools can also contribute to an already worrying democratic decline through misuse and malicious use. Proactive measures are needed to support the former and discourage the latter. While many AI

applications are unlikely to be particularly problematic, there are legitimate and pressing concerns about both malicious uses of AI and accidental or unintended harms, some of which we discuss below.

Progress in the field of computer vision has led to a rapid increase in facial recognition technologies. In 2019, the National Institute of Standards and Technology conducted a study of 189 facial recognition algorithms from 99 major tech companies and surveillance contractors. They found that a significant portion of the algorithms were biased along lines of age, race and ethnicity (Harwell 2019). Algorithms currently in use end up misidentifying some people drastically more often than others, leading to inequitable outcomes and inconsistent applications (or misapplications) of the law. Minority populations are particularly vulnerable to such flaws.

Concerningly, companies often conceal their biased algorithms under the shield of proprietary information. Biases cannot be rooted out if technologists cannot probe these algorithms for bias in the first place. Researchers at MIT compared companies which publicly named and disclosed performance results of biased AI systems to those which kept them proprietary; the former reduced classification bias in commercial APIs, whereas the latter did not (Raji and Buolamwini 2019). Promises about AI systems increasingly lean towards hype rather than science: gait recognition, eye tracking and crowd analysis make inaccurate predictions with very little oversight or public scrutiny. Scientifically questionable use cases have proliferated quickly, but remain unchallenged given the difficulty in accessing algorithms and training sets, and the lack of standardised testing. Biased algorithms and AI tools are dangerous, because they provide inaccurate information about the world. But, even more importantly, the inaccurate information they provide has real, material consequences in people's lives. In fact, there are a growing number of cases of false accusations and wrongful arrests based on the use of facial recognition systems, which remain largely unregulated.

Recent technological developments, such as improvements in generalised adversarial networks, have also led to a proliferation of deep fakes, manipulated videos and targeted propaganda. Experts predict those who seek to undermine trust in the democratic process will take advantage of an improved capacity to analyse human behaviours, moods and beliefs based on available data. According to the Center for Security and Emerging Technology, the risks include the blurring lines between foreign and domestic disinformation operations, and a widening conflict over where to draw the line between harmful disinformation and protected speech (Sedova et al 2021). Without more oversight, most technological capabilities will likely become accessible to more and more people and non-state actors. This proliferation is already ongoing in the case of spyware-for-hire companies.

Indeed, as many have noted, the challenge for researchers and governments is the dual-use potential of AI models for military and civilian purposes, for good and bad. Whereas states were typically the main actors with a capacity to influence nuclear proliferation, the development of AI applications is far more accessible to individuals and organisations. AI algorithms are often easily available and compute resources to run them are now sold as a service. Foundation models, which are models trained on broad data at scale, such that they can be adapted to a wide range of downstream tasks, are growing rapidly. Their growth will no doubt give rise to many useful and important applications, while coming with important downsides. Such models easily create or

sustain a single point of failure, meaning that vulnerabilities and biases will be blindly inherited by all the downstream uses (Bommasani et al. 2021).

3.3 Failure to invest sufficiently in AI safety and security

In addition to addressing harmful uses of AI, there is a need to accelerate the reduction of risks associated with machine learning itself. The concept of 'black box' AI, referring to the opacity of AI-based systems and how they reach particular results or predictions, is now commonplace in technology policy circles. The field of AI safety is currently grappling with a number of other technical and normative questions that are crucial to wider and safer adoption of machine learning systems.

Robustness is a key consideration. Frequently, an AI system will perform well with test data, but once deployed on a new dataset it will behave very differently. Privacy concerns are also a challenge: the deployment of an algorithm can expose sensitive information that can be gleaned directly from the model even without the release of training data. Indeed, machine learning models can be vulnerable to many different types of adversarial attacks.

An often underappreciated risk includes the difficulty of aligning human values with artificial agents (Rudner and Toner 2021) and how to deal with unexpected behaviours from artificial agents. Stuart Russell explains how AI is very good at achieving its objective: solving a puzzle, writing text that sounds human-like or winning a game. If the system finds a way to achieve the objective, it will run with it. But it is much harder to specify this objective accurately, leading to disproportionate risks of collateral. If the task of a self-driving car is to get from A to B as fast as possible, but there is no explicit consideration for sparing pedestrians on the way, accidents may well occur. The faster AI capabilities progress, the greater the risks (Russell 2020).

Governments can address these critical risks by setting benchmark standards for things like assurance and investing in potential solutions, such as privacy-preserving machine learning and 'small data' research. Even after putting laws on the books, the government must leverage its institutional and regulatory capacity to effectively enforce these standards. Despite recognition of its general importance, measures of effective AI progress are lacking from the public interest side. As general techniques, models and AI systems continue to improve and scale, policymakers and regulators need a better vocabulary and framework to quantify, understand and measure the rate of development in AI (Whittlestone and Clark 2021). This will enable more precise policy targets to be set, for example, shared measures of computational efficiency that facilitates energy saving.

These and other unresolved questions are becoming increasingly important as AI systems continue to proliferate and operate with greater autonomy at speeds that prevent humans from adequately evaluating the appropriateness of actions undertaken by machines.

3.4 Risks of market concentration and weak competition

AI innovation has been driven by twin market and research revolutions. The costs associated with AI research undoubtedly influence who can conduct such research and what outcomes they will focus their investment on. The AI renaissance emerged from the 2012 hype around reducing cost,

increasing potential, and higher feasibility of deployment for machine learning and deep learning architectures. This feasibility emerged not simply by overcoming key research challenges, but with parallel innovations across data management, data storage, parallel processing, deployment environments, open access code, open available data sets, cloud-based compute power access and related areas. As a result of these innovations, the diffusion of knowledge and the feasibility of experimentation has drastically increased.

Unlocking and driving the mass accessibility of AI, however, is increasingly led by major technology firms looking to cement leadership as the platforms and marketplaces where different AI offerings can be provided. Major firms are increasingly focused on competing over and improving the key platforms (AWS, TensorFlow), markets (app stores) and interfaces (Alexa, Siri, etc.) by which AI will be diffused globally, and long-run bets on transformative AI solutions (AGI research at OpenAI, DeepMind and Anthropic). While these are positive developments, they remain limited to a handful of firms.

The dynamism between large incumbents is insufficient to deliver a sufficiently competitive AI ecosystem (Mazzucato et al 2020). Indeed, the private sector is never incentivised to fund and develop longer-term innovations that may be globally beneficial, but are commercially risky. However, recent large-scale investments, such as Microsoft's investment in OpenAI, position the concern less as a global research contribution and more as a wager on the need to secure advantages in generalisable AI capabilities. Academic institutions, on the other hand, are in principle better placed to do this were it not for the prohibitive costs of the technological infrastructure (such as compute power) and the brain drain to the private sector (especially into technology companies such as Google, Microsoft and Facebook). Indeed, not only are academics struggling to keep pace given the high costs of infrastructure, but they're losing some of their top talent to in-house AI research in large corporations. Over time, this impacts the AI development pipeline, as academics and small startups are disadvantaged (Gofman and Jin 2022).

In the US, the National AI Research Resource Task Force is an encouraging first step that aims to spur and democratise AI-centered research and applications by developing a national cloud for scientists and students to use. Ideally, expensive experiments or datasets could be available to a wider range of institutions and researchers than is currently the case. How this is implemented, however, remains to be seen: it is important that this research cloud does not simply end up empowering large cloud companies without democratising access to cloud resources.

4. Creating and shaping AI markets

Traditionally, the role of the state in society is tied to fixing market failures, making it intrinsically reactive. However, the history of innovation policy shows that fixing market failures provides limited policy guidance for advancing the rate and direction of innovation. Market failures will no doubt continue to need fixing through the usual mechanisms, but a static state is unlikely to be future-proof, and may fail to shape the innovation system to incentivise value creation and discourage value extraction.

In recent years, the tech ecosystem has become increasingly interested in moonshots, ambitious technological projects aimed at solving some of society's most difficult challenges. While moonshots developed through the X-Prize or Google-X may push the boundaries of science, they must be aligned to the public interest to substantively address the social, political, infrastructural and economic challenges of our time. AI needs more than technological moonshots, it needs holistic missions combined with strong incentives, as well as governance structures that take responsibility for evaluating the socio-economic impacts, systemic risks and unintended consequences.

4.1 From fixing to creating and shaping markets

Former President Obama notably organised his administration's AI policy through an innovation diversity approach, 'letting a thousand flowers bloom.' By improving the overall scope and diversity of AI innovation offerings, a greater number of potential technological trajectories and innovation pathways can be explored, allowing markets to select from better technological choices, and avoid inferior innovation pathways and lock-in. However, the determinants of AI innovation go beyond any given firm or set of firms and include related institutions, processes, actors and networks that shape how R&D is organised. Understanding AI innovation capabilities as a wider innovation system helps to contextualise the design of new algorithms and the diffusion success of different programmes.

The state can play an important role again. Such goals, however, are not merely a question of aligning state capabilities with new growth objectives, but of determining the direction of innovation and the kinds of institutions and technologies which drive it. For instance, the future of smart cities is not merely a function of how next-generation digital infrastructure is set. It is also about the kinds of contracts and governance agreements shaping data collection and access, API availability, ownership, and the distribution of value from such data and opportunities. There are serious questions about liability and accountability to be answered. Alternative institutional models can radically reprioritise the kinds of infrastructure plans put in place and the distribution of benefits.

How can innovation, industrial and competition policy intersect to help shape AI ecosystems and markets towards both delivering public value and minimising market failures? What kind of policies can best accentuate the benefits of AI, while balancing or mitigating its potential harms? How can we incentivize the creation of practices and features which drive value creation, while discouraging value extraction and anti-competitive market behaviour?

4.2 Focal points for market shaping

4.2.1 Improving regulatory capabilities

AI has come under increasing scrutiny from the US Congress and the broader federal government. New ethics rules have emerged across departments, but how AI should be understood and regulated within specific sectors remains elusive for the most part. Regulating and overseeing machine learning systems is particularly difficult in light of the general-purpose

nature of the technology; the same technique (e.g. computer vision, NLP) or application (e.g. recommender systems) can be used in a variety of contexts and for different purposes. Some AI products also have dynamic and even self-modifying designs, complicating traditional 'approval-based' product regulation (Johnson 2021).

As Elen Stokes previously argued in relation to nanotechnology, new technologies and products will often confront systems of 'inherited regulation', which frequently fail to account for the nuances of new technologies. She explains that, 'Not only can it entail the application of ill-suited rules and standards, but it can also involve the reproduction of deeply ingrained traditions and assumptions which, under the weight of history, makes scrutiny extremely difficult' (2012). Adequately understanding, monitoring, evaluating and regulating AI therefore requires upskilling agencies and regulatory bodies across the board. In the United Kingdom, the Ada Lovelace Institute outlined the need for improved regulatory capacity in a recent report, noting that, 'AI systems are often complex, opaque and straddle regulatory remits,' and that, 'For the regulatory system to be able to deal with these challenges, significant improvements will need to be made to regulatory capacity' (Ada Lovelace Institute 2022). Investment into the regulatory ecosystem also means a better ability to forecast technological progress and harms, and better tools to effectively affect directionality.

The state can set regulatory outcomes, building capabilities in both internal regulators, and an intermediary market to evaluate and certify these organisations. The same logic might apply to problems regulators want to solve: Gillian Hadfield and Jack Clark's 'regulatory markets' adopts a similar logic, suggesting the creation of an intermediary layer of licensed private sector companies that compete to achieve regulatory outcomes set by a government regulator (Clark and Hadfield 2019). In other words, the objective of such a proposal is to create incentives for the private sector to allocate money, talent and computing power towards policy aims — another manifestation of directionality.

4.2.2 Directing finance

The type and quality of finance matters for driving innovation. Not all financial actors intervene at the same stage, take the same risks or invest with the same timeline of returns. Long-term patient capital is needed for transformative investments, particularly for infrastructure.

The state can act as an effective demand-side agent, working as a lender or buyer of first resort to create reliable consumers and build new market capacities around desired technological trajectories. Matt Clifford notes that, 'The Department of Defense allowed DARPA to bridge the gap between basic research and commercial application by providing real-world demand ahead of the private sector's willingness to pay' (Clifford 2022). This essential market creation capability was fundamental to accelerating and organising the direction of innovation — for example, the federal government's mass acquisition of transistors in the 1960s to accelerate NASA's moon landing mission. The lessons from this can be readily applied, for instance, to the intersection between AI and the future of green US markets, particularly with existing institutions like ARPA-E.

This model is increasingly used in innovative economies, such as the Pay for Success scheme in Taiwan. Under such an arrangement, instead of the usual cash-upfront method by which most

outsourced public services are funded, payment is instead linked to the successful achievement of a stated outcome. Taiwan's Digital Minister Audrey Tang explains that, 'An independent board assesses whether a project has delivered some return on investment in the social sense or in the environmental sense, and by the end of that evaluation period, the government is committed to pay out in a form of awards to the work that's already done' (Wiblin and Harris 2022). Importantly, this 'promise' or government commitment can then be used by startups as a security or a signal of legitimacy.

Modern US technological innovation occurs in a highly financialised system (Mazzucato 2018; Lazonick 2013; Fernandez et al 2020). A financialised system is one where financial narratives, measurements and rationalities come to dominate the organisational and investment logic of non-financial firms. William Lazonick provides a history of the role of financialisation in the US corporation, noting employment trends may be under-explained by technological change in the productive structure of the economy, and instead derive from a change in how organisations invest in management and employment, which reinforces a negative relationship on aggregative innovation performance. As a result, investing in AI firms in a financialised economy without accounting for whether a given firm or AI-driven business model drives value creation or, conversely, extraction can cause harm to the innovation system.

The investment portfolio of public actors needs to evaluate the existing conditionalities in public sector procurement contracts, such that public return on investment can be secured (Laplane and Mazzucato 2019). More fundamentally, investment and procurement contracting arrangements need to disincentivise value-extractive behaviour at the corporate governance level in associated firms to ensure that AI policy does not drive value-extractive practices in the US innovative ecosystem (Mazzucato 2013), where the private sector uses public sector money to downsize and divest from its workforce and innovation streams. Instead, concern needs to focus on how the public and private sector can create a symbiotic innovation system, where investment and partnership drive increased reinvestment into workforce and R&D instead of share buybacks and financialisation (Ibid; Mazzucato 2021).

4.2.3 Improving public sector capabilities

The US public sector has been hollowed out, reducing its capacity to effectively assess or respond to growing challenges. This trend is exacerbated by excessive outsourcing: while reliance on markets to deliver government policies is generally desirable, an overreliance on procurement and outsourcing has deprived government agencies of much-needed needed talent, capacity and expertise. While outsourcing is not problematic per se, the combination of the opacity of proprietary information and the concentration of expertise in the private sector has meant that public servants have fewer opportunities to learn and improve. Instead, they tend towards a more generalist skill set and become dependent on outsourced functions. This is a vicious cycle, as reliance on outsourced functions accelerates under-investment in capability. Under-investment leads to under-performing public services and under-equipped regulatory agents. As the National Center of Artificial Intelligence notes, the talent deficit in the Department of Defense (DOD) and the intelligence community represents the greatest impediment to being AI-ready by 2025. Talent deficits extend beyond key data science talent in to related talent for deploying and maintaining

machine learning systems, as well as integrating related innovation into core products and services, such as user experience (UX) design.

Alternative contract and grant models of employment have been used to supplement the existing capacity issues within the public sector, particularly for technological innovation and deployment. Further cause for concern is the accountability, responsibility and transparency conditions with outsourcing AI development and usage. This broad outsourcing has led to a large intermediary market of consulting agencies, notably driven by Accenture, Microsoft, IBM, Deloitte, McKinsey, BCG, Amazon and PwC — despite the trend that in-house AI development and expertise is frequently better suited to public sector tasks than outsourced models. Governments should explore significant reforms to existing procurement mechanisms. The overly complex and slow-moving nature of existing procurement frameworks should be another important target, simplifying rules for SMEs and innovators while strengthening accountability and the public interest for society more widely.

The ability to govern AI, innovate AI for the public sector, deploy and use AI in the public sector, and review, avoid or critique potential public sector AI models depends on the state of dynamic public capabilities. This likewise risks an algorithmic bureaucracy, leveraging automation as 'technological outsourcing' to further reduce public sector capabilities to achieve efficiency, while crippling government capabilities to respond to challenges or create public value. Public sector cybersecurity preparation also remains systematically underdeveloped, even without reference to more sophisticated malicious cyber behaviour; Jason Matheny estimates that only around 1% of AI research spending goes to security (Buchanan 2019). Improving public sector capabilities also implies investing more in cybersecurity.

The standards and contracting models for government AI and IT need to expand and establish a broader AI for public value innovation system, bringing together a wider cross-section of actors. However, the focus must be on internal talent and training improvement programmes, creating a long-term government learning and development process for generating new technical capabilities. Without this investment, the outsourcing and deployment of models may lead to slower service provision, underdeveloped public service offerings, and reduced efficiency due to mismatched expectations and working approaches.

The recent US Government Accountability Office proposal for a federal academy for digital staff highlights internal structural issues, most notably concerning compensation and hiring process length, and aligns with further calls for a DOD AI talent pipeline (US Government Accountability Office 2021). However, there is an urgent need to assemble an aligned programme for public sector improvement which addresses contracting models, outsourcing incentives, procurement conditions, the AI partner ecosystem, compensation and curriculum design. More generally, the civil service and government agencies are in drastic need of organisational reform and overhaul. As the necessity for governance and state involvement around AI increases, so does the need for institutions that are nimble, attract talent and deliver results. Achieving this is impossible without ambitious reform of state machinery.

5. A proactive global technology policy agenda

Since 2016, more than 60 countries have created and published AI strategies (see the OECD AI Policy Observatory for a complete and updated list of national strategies). While the US has been the core driver of AI innovation historically, its ability to create and sustain advantages derives from its domestic networks, its position in global markets and international talent flows — not, notably, from any government policy agenda or cohesive national vision. Major US corporations, such as Amazon, Apple, Google and Meta, are leading both the domestic development of AI, as well as the organisation of global access to key features for AI research and deployment, from computing power and training environments to open-sourcing algorithmic innovation. In the UK, the publicly funded Digital Catapult, a key enabler of machine learning uptake in the UK, relies on a network of corporate partners, many of which are US-based tech firms, to subsidise access to computing time.

The shape of the US AI innovation system is increasingly determining the rate and direction of AI development across Europe, Latin America, Africa and Asia. However, the US 'laissez-faire' model of AI development and usage (or lack thereof) is being pursued in parallel to other models of data collection, AI deployment and broader internet governance, most notably in Russia and China. These alternative models bring both a national security concern and a global commercial concern. The security concerns around Huawei's 5G offering, Russia's cybersurveillance ambitions and China's state-centric New IP proposal are salient examples of the growing policy proactivity of autocratic actors.

The international dimension of digital markets means that an algorithm used in one country might not present the same risks in a different country. Like-minded governments should therefore collaborate on defining a common framework for assessing and limiting both the exports and imports of potentially dangerous dual-use systems. There are many cases where commercial incentives and the availability of unscrupulous purchasers might favour the development of harmful tools, or tools used to facilitate the exercise of illiberal, illegal and harmful authoritarian practices around the world, from disinformation to spyware and censorship.

These developments — and how the US responds to them — will invariably affect key components of the AI innovation ecosystem, such as talent flows, cross-border data-sharing, foreign direct investment, international standards and cybersecurity risks. As a result, many key features of AI innovation are becoming 'geopoliticised'. For instance, the creation of the US-EU Trade and Technology Council poses a hallmark initiative for transatlantic engagement on digital innovation and development. But this opportunity also highlights two concerns: the first is the risk from an overly protectionist approach to the AI industry in US policy; the second is the future of US AI policy without a coherent vision for a future global democratic digital order.

The US needs to better articulate its vision for the future of the global digital economy, aligning its existing allies across new technology frontiers. Advancing norms of safety, privacy, accountability and transparency across global fora and rulebooks will require alignment with other democratic nations, a clear agenda and vision, and a proactive rather than reactive approach to AI policy. The EU is currently in the process of adopting the AI Act, which will propose a risk-based approach to AI that the US could seek to align with. The US and its like-minded countries should bolster and

scale initiatives like the Open Technology Fund to support and create freedom-enhancing applications of technology, helping secure the wider internet infrastructure, prevent internet shutdowns and preserving rights through technological solutions such as privacy-preserving machine learning. Inevitably, this will require the development of a clear approach to AI governance and regulation nationally. Given America's leadership in tech entrepreneurship, its domestic AI policy will necessarily shape the global market for innovation. Whatever direction the US takes will shape how entrepreneurs and firms around the world are orienting their careers and capital-raising strategies.

In addition, questions relating to AI safety will need closer scrutiny and investment. It remains unclear what competition among generalised intelligence-driven organisations will look like, as well as how this will impact the operation and development of public value-generating services. Many questions remain about how transformative AI will affect labour markets, growth, the allocation of wealth and society more widely, but a growing literature warns of potential catastrophic risks and control loss (Rudner and Toner 2021). Far from being science-fiction themes, these very well-researched risks deserve further research and decisive international policy action. Indeed, they deserve to be treated as seriously as environmental risks and should thus be addressed through coordinated global agreements.

Any US AI strategy will be fundamentally incomplete without a vision of how technology fits within the global order in which democratic and value-creating AI governance models dominate over value-extracting and autocracy-reinforcing business models. Such an order must address how the risks and rewards of AI innovation are shared, how AI can be governed and procured to address societal challenges, and how the risks and benefits of AI for economic and non-economic outcomes will be distributed. This will be the fundamental challenge shaping the next generation of global alignment on digital economic development.

6. Conclusion and recommendations

The US has a fundamental opportunity to shape the rate and direction of AI development, both domestically and in the global AI innovation system.

Technological competition and national security-based frameworks alone will fail to reflect the nature of diverging alternative futures for AI. Likewise, focusing on the initial diversity of solutions without improving the kind of institutional and infrastructural environments which avoid lock-in and path dependencies can lead to under-performing innovation systems. The US should not merely be aiming for leadership, whether in basic or applied research. The US should be aiming to build the most beneficial AI future for all within a global digital economic order.

The prior history of mission-oriented, bold investments and institutions in the US holds promise for the future of AI leadership and directionality. DARPA has led US investments since the 1950s, leveraging a defence mission and reliable buying power to create high-risk, high-reward investments. Recent National Science Foundation grants have expanded the mission-oriented approach to new research centres with additional funding for centres approved in 2021. The broadening of the mission-oriented funding scope has brought an increasing need for

coordination, leading to the establishment of the National Artificial Intelligence Office. However, the US currently lacks a model for better integrating and procuring AI, as well as adequate public sector capabilities to accelerate and democratise AI research and development.

Intelligently shaping the direction of AI research and development first requires a number of challenges to be addressed. The first is the need for governments to be able to understand the capabilities and impacts of deployed systems and ongoing AI systems more systematically. The second is access to data: many public sector organisations hold valuable datasets, but these remain inaccessible for the benefit of the wider public. Equally, the computing power required for cutting edge AI research is only available to a handful of large companies; while the price of computing has been decreasing, more can be done to facilitate research access to this type of infrastructure.

Recent technological developments, such as improvements in foundation models, have led to a host of impressive applications, but they also enable more uncertain or dangerous use cases. Governments should therefore explore ways to discourage harmful and malicious uses of AI, while proactively supporting research into safety and security. Indeed, this decade's rise in cybercrime and ransomware is a prescient warning of what policy inaction can cause. Lastly, an effective and healthy AI ecosystem necessarily means a competitive one. SMEs and researchers must gain better access to technological infrastructure, and stakeholders across the AI ecosystem must develop more programmes to attract and retain international talent.

Implementing and overseeing these interventions requires strong and democratic state capacity. The government must invest in regulatory and public sector bodies across the board, and equip them with better tools. In particular, skills and technical literacy matter greatly: internal talent and training improvement programmes are critical, as are opportunities for private sector professionals to contribute to the public sector's missions. Financing of R&D will also benefit from an overhaul. In some circumstances, the state should be willing to guarantee purchases in order to bridge the gap between basic research and market deployment, just as it recently did with vaccines in Operation Warp Speed.

Solving the above challenges will allow governments to start considering the directionality of digital transitions, and address AI risks and opportunities more systematically. In light of the fast progress we are witnessing in machine learning research, there is no better time for the US government to rethink its approach beyond global leadership in AI development towards shaping the kinds of digital economies that can be scaled to improve global public value creation.

References

Acemoglu, D. (2020). The Post-COVID State. *Project Syndicate*. Available at: https://www.project-syndicate.org/onpoint/four-possible-trajectories-after-covid19-daron-acemoglu-2020-06.

Acemoglu, D. (2021). Remaking the post-COVID world. *Finance & Development.* International Monetary Fund. Available at: https://www.imf.org/external/pubs/ft/fandd/2021/03/COVID-inequality-and-automation-acemoglu.htm.

Ada Lovelace Institute. (2022). Regulate to Innovate. Available at: https://www.adalovelaceinstitute.org/report/regulate-innovate/ (Accessed 13 June 2022).

Andreoni, A. and Roberts, S. (2020). *Governing data and digital platforms in middle income countries: Regulations, competition and industrial policies, with sectoral case studies from South Africa.* Digital Pathways at Oxford Paper Series, no. 5. Oxford, United Kingdom.

Arthur, W. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal,* 99:116-31.

Bates, J. (2013). The Strategic Importance of Information Policy for the Contemporary Neoliberal State: The Case of Open Government Data in the United Kingdom. *Government Information Quarterly* 31. 388–95.

Block, F.L. and Keller, M.R. (2011). *State of Innovation: The US Government's Role in Technology Development* (1st ed.). Routledge. DOI: https://doi.org/10.4324/9781315631905.

Bommasani, R. et al. (2021). On the Opportunities and Risks of Foundation Models. arXiv.org. Available at: https://arxiv.org/abs/2108.07258 (Accessed 13 June 2022).

Bozeman, B. (2002). Public-value failure: When efficient markets may not do. *Public Administration Review*, 62(2), pp. 145–161. DOI: https://doi.org/10.1111/0033-3352.00165.

Bozeman, B. (1988). Exploring the limits of public and private sectors: Sector boundaries as Maginot Line. *Public Administration Review*, 48(2), pp. 672. DOI: https://doi.org/10.2307/975772.

Bozeman, B. and Sarewitz, D. (2005). Public values and public failure in US science policy. *Science and Public Policy*, 32(2), pp. 119–136. DOI: https://doi.org/10.3152/147154305781779588.

Breznitz, D., Ornston, D. and Samford, S. (2018). Mission critical: The ends, means, and design of innovation agencies. *Industrial and Corporate Change*, 27(5), pp. 883–896. DOI: 10.1093/icc/dty027.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. Retrieved from: http://infolab.stanford.edu/~backrub/google.html.

Brynjolfsson, E. (2022). The Turing Trap: The promise and peril of human-like artificial intelligence. Daedalus, 151(2), pp.272–287. DOI:10.1162/daed_a_01915.

Bryson, J. (2022) Europe Is in Danger of Using the Wrong Definition of AI. Wired. Available at: https://www.wired.com/story/artificial-intelligence-regulation-european-union/.

Buchanan, B. (2019). The Future of AI and Cybersecurity. The Cipher Brief. Available at: https://www.thecipherbrief.com/column_article/the-future-of-ai-and-cybersecurity (Accessed 13 June 2022).

Clark, J. and Hadfield, G. (2019). Regulatory Markets for AI Safety. arXiv.org. Available at: https://arxiv.org/abs/2001.00078 (Accessed 13 June 2022).

Clifford , M. (2022). Dominic Cummings got his British Darpa. Can he make it work? Wired UK. Available at: https://www.wired.co.uk/article/dominic-cummings-british-darpa (Accessed 13 June 2022).

Collington, R. (2022). Disrupting the Welfare State? Digitalisation and the Retrenchment of Public Sector Capacity. *New Political Economy*, 27:2, 312-328, DOI: 10.1080/13563467.2021.1952559

Culpepper, P. D. and Thelen, K. (2019). Are we all Amazon Primed? Consumers and the politics of platform power. *Comparative Political Studies*, 53(2), pp. 288–318. DOI: https://doi.org/10.1177/0010414019852687.

Dattani, S. and Bechhofer, N. (2021). The speed of science. Works in Progress. Available at: https://www.worksinprogress.co/issue/the-speed-of-science/ (Accessed 13 June 2022).

Deloitte. (2017). Assessing the value of TfL's open data and digital partnerships. Available at: https://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf.

DeVries, T., Misra, I., Wang, C. and van der Maaten, L. (2019). Does Object Recognition Work for Everyone? Meta Research. Available at: https://research.facebook.com/publications/does-object-recognition-work-for-everyone/ (Accessed 13 June 2022).

Dewey, J. (1954). *The Public and its Problems*. Athens, OH: Swallow.

Fernandez, R.., Adriaans, I., Hendrikse, R. and Klinge, T.J. (2020). The financialisation of Big Tech. SOMO. Available at: https://www.somo.nl/the-financialisation-of-big-tech/ (Accessed 13 June 2022).

Gofman, M. and Jin, Z. (2022). Artificial Intelligence, Education and Entrepreneurship. Available at https://ssrn.com/abstract=3449440. DOI: http://dx.doi.org/10.2139/ssrn.3449440.

Harwell, D. (2019). Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use. Washington Post. Available at: https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/ (Accessed 13 June 2022).

Hogarth, I. and Benaich, N. (2021). *State of AI Report 2021*. Available at: https://www.stateof.ai/

--(2020) *State of AI report 2020*. Available at: https://www.stateof.ai/2020

--(2019) *State of AI report 2019*. Available at: https://www.stateof.ai/2019

--(2018) *State of AI report 2018*. Available at: https://www.stateof.ai/2018

Johnson, W. G. (2021). Flexible regulation for dynamic products? The case of applying principles-based regulation to medical products using artificial intelligence. *Law, Innovation and Technology*, 14(2). Available at SSRN: https://ssrn.com/abstract=3850968. DOI: http://dx.doi.org/10.2139/ssrn.3850968.

Karo, E. and Kattel, R. (2018). Innovation and the State: Towards an Evolutionary Theory of Policy Capacity. In: Wu X., Howlett M. and Ramesh M. (eds.), *Policy Capacity and Governance: Assessing Governmental Competences and Capabilities in Theory and Practice*. Studies in the Political Economy of Public Policy. Cham: Springer International Publishing, pp. 123–150. DOI: 10.1007/978-3-319-54675-9_6.

Kattel, R., Drechsler, W. and Karo, E. (2019). *Innovation bureaucracies: How agile stability creates the entrepreneurial state*. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2019-12). Available at: https://www.ucl.ac.uk/bartlett/public-purpose/wp2019-12.

Kattel, R. and Mazzucato, M. (2018a). *Mission-oriented innovation policy and dynamic capabilities in the public sector*. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2018-5). Available at: https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/iipp-wp-2018-05.pdf.

Keynes, J. M. (1942). *The General Theory of Employment, Interest and Money*. Macmillan.

Khan, L. (2017). Amazon's antitrust paradox. *Yale Law Journal*, 126, pp. 710-805

Khan, L. (2018). The New Brandeis Movement: America's antimonopoly debate. *Journal of European Competition Law & Practice*, 9(3), pp. 131–132. DOI: https://doi.org/10.1093/jeclap/lpy020.

Laplane, A. and Mazzucato, M. (2019). *Socialising the risks and rewards of public investments: Economic, policy and legal issues*. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2019-09). Available at: https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/socialising_risks_and_rewards_final.pdf.

Lazonick, W. (2014). Profits Without Prosperity. *Harvard Business Review*, September.

Lazonick, W. and Mazzucato, M. (2013). The risk-reward nexus in the innovation-inequality relationship: who takes the risks? Who gets the rewards? *Industrial and Corporate Change*, 22(4), 1093-1128.

Mazzucato, M. (2013). *The Entrepreneurial State: Debunking the Public vs. Private Myths in Risk and Innovation*. Anthem Press

Mazzucato, M. (2016). From Market Fixing to Market-Creating: A new framework for innovation policy. *Industry and Innovation*, 23(2), 140-156.

Mazzucato, M. (2018). Mission Oriented Innovation Policy: Challenges and Opportunities. *Industrial and Corporate Change*, 27 (5): 803–815.

Mazzucato, M. (2021). *Mission Economy. A Moonshot Guide to Changing Capitalism*. Allen Lane.

Mazzucato, M., Andreoni, A. and Conway, R. (2021). *Mission-oriented innovation in the U.S.A: Shaping markets toward grand challenges: A new industrial policy frontier*. UCL Institute for Innovation and Public Purpose, Policy Note No. 001.

Mazzucato, M., Entsminger, J. and Kattel, R. (2020). *Public value and platform governance.* UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2020-11). Available at: https://www.ucl.ac.uk/bartlett/public-purpose/wp2020-11.

Mazzucato, M. and Kattel, R. (2020). COVID-19 and public sector capacity. *Oxford Review of Economic Policy*.

Mazzucato, M., Kattel, R. and Ryan-Collins, J. (2020). Challenge-driven innovation policy: Towards a new policy toolkit. *Journal of Industry, Competition and Trade,* 20, 421–437. DOI: https://doi.org/10.1007/s10842-019-00329-w.

Mazzucato, M. and Ryan-Collins, J. (2019). Putting value creation back into 'public value': From market fixing to market shaping. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2019-05).

Mazzucato, M., Ryan-Collins, J. and Giorgos, D. (2020). Theorising and mapping modern economic rents. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2020-13). Available at: https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/final_iipp-wp2020-13-theorising-and-mapping-modern-economic-rents_8_oct.pdf.

Morozov, E. (2013). *To Save Everything, Click Here: Technology, Solutionism and the Urge to Fix Problems that Don't Exist*. London: Allen Lane.

Nelson, R. (1990). US technological leadership: Where did it come from and where did it go? *Research Policy*, 19:117-132.

National Research Council. (1999). *Funding a Revolution: Government Support for Computing Research*. Washington, DC: The National Academies Press. DOI: https://doi.org/10.17226/6323.

NSCAI. (2021). National Security Commission on Artificial Intelligence Final Report. Available at: https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

OECD AI Policy Observatory. Available at: https://oecd.ai/en/.

Perez, C. (2003). *Technological Revolutions and Financial Capital*. Cheltenham, UK: Edward Elgar Publishing.

Peters, B. G. (2017). What is so wicked about wicked problems? A conceptual analysis and a research program. *Policy and Society*, 36(3), pp. 385–396. DOI: https://doi.org/10.1080/14494035.2017.1361633.

Polanyi, K. (2001). (1944). *The Great Transformation: The Political and Economic Origins of Our Time*. (2nd Beacon Paperback ed.). Beacon Press, Boston.

Quantum Commodity Intelligence. (2022). IEA mulls whether to make all its data freely available. Available at: https://www.qcintel.com/article/correction-iea-proposes-to-make-all-its-data-freely-available-3540.html (Accessed 13 June 2022).

Raji, I. and Buolamwini, J. (2019). Actionable auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics and Society, pp. 429–435. DOI: https://doi.org/10.1145/3306618.3314244.

Rikap, C. (2021). *Capitalism, Power and Innovation: Intellectual Monopoly Capitalism Uncovered* (1st ed.). Routledge. DOI: https://doi.org/10.4324/9780429341489.

Rodrik, D. (2014). Green industrial policy. *Oxford Review of Economic Policy*, 30(3): 469–491.

Rudner, T. G. J. and Toner, H. (2021). Key Concepts in AI Safety: An Overview. Center for Security and Emerging Technology. Available at: https://cset.georgetown.edu/publication/key-concepts-in-ai-safety-an-overview/ (Accessed 13 June 2022).

Russell, S. (2020). Human Compatible: AI and the Problem of Control. Penguin.

Sedova, K., McNeill, C., Johnson, A., Joshi, A. and Wulkan, I. (2021). AI and the Future of Disinformation Campaigns. Center for Security and Emerging Technology. Available at: https://cset.georgetown.edu/publication/ai-and-the-future-of-disinformation-campaigns/ (Accessed 13 June 2022).

Stanford Digital Economy Lab. (2022). The Turing Trap: The promise and peril of human-Like artificial intelligence. Available at: https://digitaleconomy.stanford.edu/news/the-turing-trap-the-promise-peril-of-human-like-artificial-intelligence/ (Accessed 13 June 2022).

Stirling, A. (2008). 'Opening up' and 'closing down': Power, participation and pluralism in the social appraisal of technology. *Science, Technology and Human Values*, *33*(2), 262–2

Stirling, A. (2015). Emancipating transformations: From controlling 'the transition' to culturing plural radical progress. In: Scoones, I., Leach, M. and Newell, P. (eds) *The Politics of Green Transformations*. Available at: https://www.taylorfrancis.com/chapters/oa-edit/10.4324/9781315747378-4/emancipating-transformations-andy-stirling (Accessed 13 June 2022).

Stokes, E. (2012). Nanotechnology and the products of inherited regulation. Journal of Law and Society, Vol. 39, Issue 1, pp. 93-112. Available at SSRN: https://ssrn.com/abstract=2009202. DOI: http://dx.doi.org/10.1111/j.1467-6478.2012.00572.x.

Teece, D. and Pisano, G. (1994). The dynamic capabilities of firms: An introduction. *Industrial and Corporate Change*, 3(3), pp. 537–556. DOI: 10.1093/icc/3.3.537-a.

Tucker, C. (2018). Network effects matter less than they used to. That's a really big deal. Retrieved from: https://hbr.org/2018/06/why-network-effects-matter-less-than-they-used-to.

Tucker, T. (2019). *Industrial Policy and Planning: What It Is and How to Do It Better.* Roosevelt Institute. Available at: https://rooseveltinstitute.org/publications/industrial-policy-and-planning/.

US Government Accountability Office. (2021). Digital Services: Considerations for a Federal Academy to Develop a Pipeline of Digital Staff. GAO. Available at: https://www.gao.gov/assets/gao-22-105388.pdf (Accessed 13 June 2022).

US OTA. (1989). *Holding the Edge: Maintaining the Defence Technology Base*. Washington, DC: Government Printing Office.

Wachter, S. and Mittelstadt, B. (2018). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, 2019(2), pp. 494–620. DOI: https://doi.org/10.7916/cblr.v2019i2.3424

Wachter, S., Mittelstadt, B. and Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, Volume 41. DOI: https://doi.org/10.1016/j.clsr.2021.105567.

Wakefield, J. (2021). AI emotion-detection software tested on Uyghurs. BBC News. Available at: https://www.bbc.co.uk/news/technology-57101248 (Accessed 13 June 2022).

Whittlestone, J. and Clark, J. (2021). Why and How Governments Should Monitor AI Development. arXiv.org. Available at: https://arxiv.org/abs/2108.12427 (Accessed 13 June 2022).

Wiblin, R. and Harris, K. (2022). Audrey Tang on what we can learn from Taiwan's experiments with how to do democracy. 80,000 Hours. Available at: https://80000hours.org/podcast/episodes/audrey-tang-what-we-can-learn-from-taiwan/ (Accessed 13 June 2022).

Zuboff, S. (2020). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

The IIPP Working Paper series and Policy Reports

IIPP WP 2017-01 Mission-Oriented innovation policy: Challenges and opportunities. Mariana Mazzucato.

IIPP WP 2017-02 Thinking about technology policy: 'Market Failures' versus 'Innovation systems'. Richard R Nelson.

IIPP WP 2017-03 Technological capacity in the public sector: the Case of Estonia. Veiko Lember, Rainer Kattel, Piret Tõnurist.

IIPP WP 2017-04 Rethinking value in health Innovation: From mystification towards prescriptions. Mariana Mazzucato, Victor Roy.

IIPP WP 2017-05 Patient strategic finance: Opportunities for state investment banks in the UK. Mariana Mazzucato, Laurie Macfarlane.

IIPP WP 2018-01 State investment banks and patient finance: An international comparison. Laurie Macfarlane, Mariana Mazzucato.

IIPP WP 2018-02 Putting austerity to bed: Technical progress, aggregate demand and the supermultiplier. Matteo Deleidi, Mariana Mazzucato.

IIPP WP 2018-03 The bit and the rainforest: Towards the evolutionary theory of policy capacity. Erkki Karo, Rainer Kattel.

IIPP WP 2018-04 Financing green growth. Semieniuk Gregor, Mariana Mazzucato.

IIPP WP 2018-05 Mission-oriented innovation policy and dynamic capabilities in the public sector. Rainer Kattel, Mariana Mazzucato.

IIPP WP 2018-06 The economics of change: Policy and appraisal for missions, market shaping and public purpose. Rainer Kattel, Mariana Mazzucato, Josh Ryan-Collins, Simon Sharpe.

IIPP WP 2018-07 Movements with missions make markets. Charles Leadbeater.

IIPP WP 2018-08 Bringing the helicopter to ground: A historical review of fiscal-monetary coordination to support economic growth in the 20th century. Josh Ryan-Collins, Frank van Lerven.

IIPP WP 2018-09 Estonia's digital transformation: Mission mystique and the hiding hand. Rainer Kattel, Ines Mergel.

IIPP WP 2018-10 The people's prescription: Re-imagining health innovation to deliver public health. UCL Institute for Innovation and Public Purpose, Stop Aids, Just Treatment, Global Justice Now.

IIPP WP 2018-11 Credit where it's due: A historical, theoretical and empirical review of credit guidance policies in the 20th century. Dirk Bezemer, Josh Ryan-Collins, Frank van Lerven and Lu Zhang.

IIPP WP 2019-01 Supermultiplier, innovation and the ecosystem: A stock-flow dynamic model. Matteo Deleidi, Riccardo Pariboni, Marco Veronese Passarella.

IIPP WP 2019-02 A mission-oriented framework for the Scottish National Investment Bank. Mariana Mazzucato, Laurie Macfarlane.

IIPP WP 2019-03 A framework for mission-oriented innovation policy roadmapping for the SDGs. Mariana Mazzucato, Laurie Macfarlane.

IIPP WP 2019-04 A Mission-Oriented UK Industrial Strategy. UCL Commission for Mission-Oriented Innovation and Industrial Strategy (MOIIS).

IIPP WP 2019-05 Putting value creation back into 'public value': From market fixing to market shaping. Mariana Mazzucato, Josh Ryan-Collins.

IIPP WP 2019-06 The macroeconomic impact of government innovation policies: A quantitative assessment. Matteo Deleidi, Vicenzo De Lipsis, Mariana Mazzucato, Josh Ryan-Collins, Paolo Agnolucci.

IIPP WP 2019-07 Financial and legal barriers to the creation and operation of a British national investment bank. Rob Calvert Jump, Natalya Naqvi.

IIPP WP 2019-08 Public investment fiscal multipliers: An empirical assessment for European countries. Matteo Deleidi, Francesca Iafrate, Enrico Sergio Levrero.

IIPP WP 2019-09 Socialising the risks and rewards of public investments: Economic, policy and legal issues. Andrea Laplane, Mariana Mazzucato.

IIPP WP 2019-10 Neither crowding in nor out: Public direct investment mobilising private investment into renewable electricity projects. Matteo Deleidi, Mariana Mazzucato, Gregor Semieniuk.

IIPP WP 2019-11 Social structures of accumulation in Greece, 1980-2014. Angeliki Papadopoulou, Giorgos Gouzoulis.

IIPP WP 2019-12 Innovation bureaucracies: How agile stability creates the entrepreneurial state. Rainer Kattel, Wolfgang Drechsler, Erkki Karo.

IIPP WP 2019-13 Climate-related financial policy in a world of radical uncertainty: Towards a precautionary approach. Hugues Chenet, Josh Ryan-Collins, Frank van Lerven.

IIPP WP 2020-01 The public economy: Understanding government as a producer. June Sekera.

IIPP WP 2020-02 The entrepreneurial (welfare) state? Tackling social issues through challenge prizes. Ville Takala, Emma Nordbäck and Tuukka Toivonen.

IIPP WP 2020-03 Determinants of income shares and the stable middle in post-socialist China. Giorgos Gouzoulis, Collin Constantine.

IIPP WP 2020-04 Industrial policy: A long-term perspective and overview of theoretical arguments Erik S. Reinert.

IIPP WP 2020-05 Gig work at the base of the pyramid: considering dependence and control. Kate Roll.

IIPP WP 2020-06 Deindustrialisation reconsidered: Structural shifts and sectoral heterogeneity. Fiona Tregenna, Antonio Andreoni.

IIPP WP 2020-07 Upward-scaling tipping cascades to meet climate goals: Plausible grounds for hope. Simon Sharpe, Timothy Lenton.

IIPP WP 2020-08 When homes earn more than jobs: the rentierization of the Australian housing market. Josh Ryan-Collins, Cameron Murray.

IIPP WP 2020-09 Managing nature-related financial risks: a precautionary policy approach for central banks and financial supervisors. Katie Kedward, Josh Ryan-Collins, Hugues Chenet.

IIPP WP 2020-10 Welfare 5.0: Why we need a social revolution and how to make it happen. Hilary Cottam.

IIPP WP 2020-11 Public value and platform governance. Mariana Mazzucato, Josh Entsminger, Rainer Kattel.

IIPP WP 2020-12 COVID-19 and public-sector capacity. Mariana Mazzucato, Rainer Kattel.

IIPP WP 2020-13 Theorising and mapping modern economic rents. Mariana Mazzucato, Josh Ryan-Collins, Giorgos Gouzoulis.

IIPP WP 2020-14 Missioni Italia. Mariana Mazzucato.

IIPP WP 2020-15 Schumpeter, the entrepreneurial State and China. Leonardo Burlamaqui.

IIPP WP 2020-16 Public wealth funds: Supporting economic recovery and sustainable growth. Dag Detter, Stefan Fölster, Josh Ryan-Collins.

IIPP WP 2020-17 The EIB and the new EU missions framework. Mariana Mazzucato, Olga Mikheeva.

IIPP WP 2020-19 Creating and measuring dynamic public value at the BBC. Mariana Mazzucato, Rowan Conway, Eleonora Maria Mazzoli, Eva Knoll, Sarah Albala.

IIPP WP 2020-20 The entrepreneurial State and public options: socialising risks and rewards. Mariana Mazzucato, Henry Lishi Li.

IIPP WP 2020-21 A market-shaping approach for the biopharmaceutical industry: governing innovation towards the public interest. Mariana Mazzucato, Henry Lishi Li.

IIPP WP 2021/01 Dynamic capabilities in the public sector: The case of the UK's Government Digital Service. Rainer Kattel, Ville Takala.

IIPP WP 2021/02 Innovations in development finance and conditioning factors: BNDES and the fostering of sustainability-related industries in Brazil. João Carlos Ferraz, Luma Ramos, Bruno Plattek.

IIPP WP 2021/03 Deciding how to decide: Risk-opportunity analysis as a generalisation of cost-benefit analysis. Michael Grubb, Matthew Ives, Florian Knobloch, Jean-Francois Mercure, Femke J. M. M. Nijsse, Hector Pollitt, Simon Sharpe, Jorge Vinuales.

IIPP WP 2021/04 New economics of assistive technologies: A call for a missions approach. Sarah Albala, Catherine Holloway, Victoria Austin, Rainer Kattel.

IIPP WP 2021/05 Why ethics of quantification is needed now. Andrea Saltelli, Antonio Andreoni, Wolfgang Drechsler, Jayati Ghosh, Rainer Kattel, Ingrid H. Kvangraven, Ismael Rafols, Erik S. Reinert, Andy Stirling, Ting Xu.

IIPP WP 2021/06 A dynamic theory of public banks (and why it matters). Thomas Marois.

IIPP WP 2021/07 Altered States: Cartesian and Ricardian dreams. Erik S. Reinert, Monica di Fiore, Andrea Saltelli, Jerome R. Ravetz.

IIPP WP 2021/08 Lessons from the past 21st century systems of state-owned enterprises: The case of Italy's IRI in the 1930s. Simone Gasperin.

IIPP WP 2021/09 Building state capacities and dynamic capabilities to drive social and economic development: The case of South Africa. Mariana Mazzucato, Mzukisi Qobo, Rainer Kattel.

IIPP WP 2021/10 Mission-oriented policies and the "Entrepreneurial State" at work: An agent-based exploration. Giovanni Dosi, Francesco Lamperti, Mariana Mazzucato, Mauro Napoletano, Andrea Roventini.

IIPP WP 2021/11 Heterogeneous investors, scale economies and the commercialisation of innovative renewable energy. Gregor Semieniuk, José Alejandro Coronado, Mariana Mazzucato.

IIPP WP 2022/01 Governing finance to support the net-zero transition: lessons from successful industrialisation. Olga Mikheeva, Josh Ryan-Collins.

IIPP WP 2022/02 A new data deal: the case of Barcelona. Fernando Monge, Sarah Barns, Rainer Kattel and Francesca Bria.

IIPP WP 2022/03 How can South Africa advance a new energy paradigm? A mission-oriented approach to megaprojects. Antonio Andreoni, Kenneth Creamer, Mariana Mazzucato, Grové Steyn.

IIPP WP 2022/04 Policy capacities for transformative innovation policy: A case study of UK Research Innovation. Julie McLaren, Rainer Kattel.

IIPP WP 2022/05 From financial risk to financial harm: Exploring the agri-finance nexus and drives or biodiversity loss. Katie Kedward and Josh Ryan-Collins.

IIPP WP 2022/07 Dynamic capabilities of the public sector: Towards a new synthesis. Rainer Kattel.

IIPP WP 2022/08 The self-financing state: An institutional analysis of government expenditure, revenue collection and debt issuance operations in the United Kingdom. Andrew Berkeley, Josh Ryan-Collins, Richard Tye, Asker Voldsgaard and Neil Wilson.

IIPP WP 2022/09 Expanding DARPA's model of innovation for biopharma: A proposed Advanced Research Projects Agency for Health. Mariana Mazzucato and Travis Whitfill.

All Working Papers are available to download at the Institute for Innovation and Public Purpose website: ucl.ac.uk/iipp

ucl.ac.uk/iipp
@IIPP_UCL

UCL Institute for Innovation and Public Purpose
11 Montague Street, London, WC1B 5BP

General enquiries:
iipp-research@ucl.ac.uk
Tel: +44 (0)20 3108 6961